SHORT REVIEW

Hox transcription factors and their elusive mammalian gene targets

T Svingen and KF Tonissen

Cell Biology Group, Eskitis Institute for Cell and Molecular Therapies and School of Biomolecular and Biomedical Science, Griffith University, Nathan, Queensland 4111, Australia

The Hox family of homeodomain transcription factors regulate numerous pathways during developmental and normal cellular processes. All Hox proteins recognise similar sequences *in vitro* yet display functional diversity in an *in vivo* environment. This review focuses on the transcriptional and functional specificity elicited by Hox proteins, giving an overview of homeodomain–DNA interactions and the gain of binding specificity through cooperative binding with cofactors. Furthermore, currently identified mammalian Hox target genes are presented, of which the most striking feature is that very few direct Hox targets have been identified. The direct targets participate in an array of cellular functions including organogenesis and cellular differentiation, cell adhesion and migration and cell cycle and apoptotic pathways. A further assessment of identified mammalian promoter targets and the contribution of bases outside the canonical recognition motif is given, highlighting roles they may play in either *trans*-activation or repression by Hox proteins.

Introduction

Homeobox genes code for transcription factors that act during development and were first identified in the fruit fly *Drosophila melanogaster* (Lewis, 1978). A common feature of these proteins is the presence of a highly conserved 60-amino-acid motif, the homeodomain, that is responsible for binding to DNA at specifically recognised binding sites resulting in transcriptional regulation of their target genes, as previously reviewed (Levine and Hoey, 1988; McGinnis and Krumlauf, 1992). At present, a large number of homeobox genes have been characterised, now comprising an extensive superfamily of regulatory genes with members found in all animal species. There is a large catalogue of divergent homeobox families, of which the best characterised and most extensively studied is the Hox family.

In mammalian species, there are 39 Hox genes organised in four clusters labelled A, B, C and D located on four different chromosomes and numbered from 1 to 13, although no cluster contains a full set. They are thought to have arisen through the combination of *cis*-amplification and *trans*-duplication of the *Drosophila bithorax* and *antennapedia* complexes during separate evolutionary events (Scott, 1992). As a consequence, the *trans*-paralogues (genes occupying the same relative position along the 5′–3′ coordinate and hence labelled with the same number) share a higher degree of sequence homology than the *cis*-paralogues (genes occupying adjacent positions along the 5′–3′ coordinate). The high degree of homology between *trans*-paralogues immediately suggested they could act as functional complements, which has been confirmed experimentally for *Hoxa3* and *Hoxd3* (Greer *et al*, 2000). However, other Hox paralogues exhibit distinct biological roles possibly due to subfunctionalisation (one copy has acquired a new function) or neofunctionalisation processes (the original gene function is split between the copies) (Walsh, 2003).

A significant research effort has been spent trying to elucidate the exact mechanisms by which the Hox proteins gain their functional specificity and where in the regulatory hierarchy they exert their effect. Even so, there are still large gaps in the puzzle that eventually will allow us to fully appreciate the intricate control mechanisms governed by the Hox proteins.

This review will summarise much of what is known regarding Hox transcriptional specificity, with the main focus on the mammalian Hox family. As a great deal of our understanding of both Hox transcriptional and functional specificity comes from studies on *Drosophila* orthologues, however, these studies will form an integral part of this paper. The homeodomain, including its functionality, is greatly conserved through evolution and therefore nonmammalian systems are invaluable towards deciphering the regulatory principles governing Hox specificity.

Transcriptional *versus* functional specificity – the homeodomain paradox

The homeodomain consists of three alpha helices that form a helix–turn–helix motif and an additional domain

known as the N-terminal arm, located just adjacent to the first helix. This homeodomain structure has been largely conserved across species, from flies to humans, whereas the remainder of the protein has not. Functional specificity has been mapped to amino-acid residues within these conserved regions of the homeo-domain (Kuziora and McGinnis, 1989; Lin and McGinnis, 1992). As Hox functional specificity has been conserved over large evolutionary distances (Malicki *et al*, 1990; McGinnis *et al*, 1990), the term Hox specificity has become somewhat synonymous with homeodomain specificity.

A core DNA consensus sequence is recognised by the majority of homeodomain proteins, raising the question as to how transcriptional and biological specificity can be obtained. The many homeodomain-containing transcription factors elicit distinct and different effects on downstream target genes, controlling unique facets within the transcriptional machinery and thus ultimately cellular fate, but through what mechanisms? Of course, DNA sequence recognition alone is not the only determinant for biological specificity, however, sequence-specific binding is a critical part of transcriptional regulation and thus biological specificity.

Binding of the homeodomain to a specific DNA sequence was described using findings from both genetic and structural studies. Genetically, a conserved TAAT-motif was identified as a high-affinity binding site for numerous homeodomain proteins (Desplan *et al*, 1988). X-ray crystallography (Kissinger *et al*, 1990) and nuclear magnetic resonance spectroscopy (Otting *et al*, 1990) techniques revealed the helix–turn–helix motif of the homeodomain folds to contact DNA in the major groove, with an additional contact in the minor groove through the N-terminal arm. Using the same convention for amino acid numbering of the homeodomain as originally described (Qian *et al*, 1989) basic residues at positions 3 and 5 bind the T_1 and A_2 nucleotides, whereas a conserved Asn at position 51 binds the nucleotide A_3 and finally a hydrophobic residue at position 47 binds the T_4 nucleotide (Kissinger *et al*, 1990) (Figure 1). Mammalian Hox proteins also display protein–DNA interactions involving N-terminal arm residue 5 (Arg) and helix 3 residues 51 (Asn) and 47 (Ile) for both Hoxb1 (Piper *et al*, 1999) and Hoxa9 (LaRonde-LeBlanc and Wolberger, 2003).

As the preferred DNA-recognition core of many homeodomain proteins is <u>TA</u>AT and the 5′ base pairs of the motif are important for selecting *in vitro* binding sites, the homeodomain appears to possess the ability to discriminate between bases in the minor groove, despite the usual expectation that minor grove interactions yield lower binding specificities due to structural and bio-chemical properties of the double helix (Seeman *et al*, 1976). Nevertheless, studies showed that homeodomains can differentiate between A:T and T:A base pairs at position 1 of the recognition core (Ades and Sauer, 1995). Ades and Sauer (1995) also found that the two base pairs immediately 3′ of the core motif (positions 5 and 6) display a greater specificity than positions 1 and 2 and implicated amino-acid residue 50 of the homeodomain as involved in contacting these bases.

Mutational studies confirmed that residue 50 of the homeodomain can alter binding specificity, with the preferred *engrailed* binding motif TAAT<u>TA</u> changing to
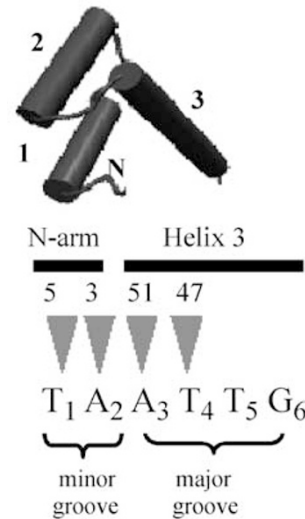
Figure 1 The homeodomain contacts the DNA recognition sequence through amino-acid residues in the N-terminal arm and third helix. A 3D view of the homeodomain is shown arranged over the consensus recognition sequence. The N-terminal arm is labelled 'N' and the three alpha helices are also indicated with their respective numbers. The contacts between the homeodomain amino-acid residues 3, 5, 47 and 51 and the TAAT binding motif are delineated by triangles. Amino acids 3 and 5 of the homeodomain N-terminal arm make contacts in the minor groove whereas amino acids 47 and 51 (in helix 3) make contact in the major groove (Qian *et al*, 1989; Piper *et al*, 1999; LaRonde-LeBlanc and Wolberger, 2003). Nucleotides 5 and 6 are also thought to make specific contact with the homeodomain through other helix 3 residues (Ades and Sauer, 1995).

TAAT<u>CC</u> when Gln50 is substituted with a Lys50 (Ades and Sauer, 1994). And surprisingly, although all Hox proteins encode a Gln at position 50, they still show some intrinsic ability to discriminate between selected base pairs 3′ of the TAAT core (Ekker *et al*, 1992). In a study of five Hox proteins (Hoxa5, Hoxa7, Hoxb1, Hoxb4 and Hoxc8), the 3′-flanking base pair (position 5) had consequences for DNA binding affinity. The sequence TAATT<u>G</u> was defined as the optimal motif and a substitution at base pair position 5 from T to G was only tolerated by Hoxb4, but not by the other four Hox proteins. The substitution at this position from T to either A or C reduced the binding activity of all five Hox proteins and further alterations in the base pair composition from position 6 onwards had varying effects on Hox binding specificity (Pellerin *et al*, 1994).

The fact that base pairs immediately 3′ of the core TAAT motif are involved in DNA binding specificity of Hox proteins is puzzling. For divergent homeodomain protein families, this may be explained by a varying amino-acid composition of helix 3, in particular residue 50. The Hox proteins, however, are highly conserved in helix 3, in particular those residues implicated in direct protein–DNA binding. All 39 Hox proteins contain the same six amino acids WFQNRR at positions 48–53, encompassing the conserved Gln50 residue. If residue 50 is responsible for dictating specificity for base pair 5, why then does Hoxb4 tolerate a G at this position, whereas Hoxa5, Hoxa7, Hoxb1 and Hoxc8 do not? The

existence of Hox cofactors reveals a further dimension of DNA binding affinity and specificity.

Hox cofactors and transcriptional specificity

Pbx1, a non-Hox homeodomain protein orthologue of the *Drosophila extradenticle* (*exd*) encoded protein, was the first characterised Hox cofactor. The cofactor interactions are highly dependent on a hexapeptide sequence containing four core residues YPWM (located just N-terminal of the Hox homeodomain), which are conserved among Hox paralogue groups 1–8 (Chang *et al*, 1995; Phelan *et al*, 1995). Hox paralogue groups 9–10, which lack the consensus YPWM motif also cooperatively bind to DNA with Pbx1, with binding dependent on a conserved Trp residue (Chang *et al*, 1995; Shen *et al*, 1997b). Another member of the TALE family of homeodomain proteins is Meis1, an orthologue of *Drosophila* Homothorax (Hth), which acts as a DNA binding partner for the remaining paralogue groups 11–13 (Shen *et al*, 1997a). Thus, all 39 Hox proteins can cooperatively bind to DNA through heterodimerisation with other non-Hox homeodomain proteins, with both preferred DNA recognition sequence and binding partner being determined by their position on the 5'–3' chromosomal coordinate.

Evidently, the Hox proteins show a relatively relaxed DNA binding selectivity *in vitro*, ultimately suggesting a more stringent selectivity *in vivo* through cooperative DNA binding with other proteins. And even though these mechanisms are not fully understood, a reasonable explanation is that cofactors may reveal latent specificities intrinsic to the individual Hox homeodomains (Mann and Morata, 2000). This would also partially explain the increase in binding affinity by Hox proteins bound to DNA in conjunction with a cofactor, compared to binding affinities of Hox monomers.

It is also noteworthy that many of the identified Hox DNA recognition sites within gene promoters are paired with Pbx/Meis binding sites, strongly suggesting the importance of the TALE-class proteins as Hox cofactors *in vivo*. Therefore, one would expect abrogated Pbx/Meis function to mimic those effects observed for mutated Hox proteins, and even mutations in the Hox hexapeptide to resemble phenotypic effects observed for Hox loss-of-function mutations. Indeed, the latter was observed with hexapeptide-targeted mutations of *Hoxb8* (Medina-Martinez and Ramirez-Solis, 2003). In this *in vivo* mouse model, Hoxb8 was prevented from interacting with its Pbx cofactor, resulting in a dominant phenotype characterised by an anterior homeotic transformation of the first two thoracic vertebrae without affecting the expression pattern of other closely related Hoxb genes. Hence, these studies put forward the view that Hoxb8 in the absence of its DNA binding partner could occupy binding elements of closely related proteins such as Hoxb6 and Hoxb7, resulting in abnormal *trans*-regulation of Hox target genes.

The above-mentioned study also highlights the importance of *in vivo* models to fully appreciate the functional specificity elicited by Hox transcription factors. Clearly, they can bind DNA elements both as monomers and cooperatively with other factors such as TALE-class proteins. Similarly, tissue-specific coexpression of Hox cofactors likely guides the Hox proteins to their appropriate gene targets, partly explaining the dilemma regarding their uniform *in vitro* recognition elements.

Hox proteins function as both transcriptional repressors and activators

Another dimension of Hox transcriptional specificity is their ability to act as both transcriptional repressors and activators and although the mechanisms behind this regulatory switch are largely unclear, some postulates have been put forward. (Pinsonneault *et al*, 1997) hypothesised a model in which heterodimerisation of the *Drosophila* Exd protein with Hox proteins abolishes the intrinsic repressor function of Hox proteins, which is exerted onto the target gene when bound as monomers, offering a simple, yet elegant model for Hox functional specificity. There are other observations contradicting this model however. More recently Exd, as well as the *Drosophila* protein Hth was shown to repress the *distalless* (*dll*) gene through direct interactions with the *Ultrabithorax* (*Ubx*) gene product (Gebelein *et al*, 2002). Hence, heterodimers may also act as suppressor complexes in this context.

While mammalian Hox transcription factors also exhibit lower DNA binding affinity when bound as monomers and thus potentially lower specificity (Shen *et al*, 1996, 1997a, b) they also contradict a model where they always act as repressors when bound weakly in monomeric form and as activators when bound with cofactors in stronger affinity complexes (Shen *et al*, 2004). The discovery that Hox proteins bind to the CREB binding protein (CBP) and can modulate histone acetyltransferase activity indicates that Hox proteins can also regulate gene expression without binding DNA directly (Shen *et al*, 2001) and further complicates the search for downstream targets.

Mammalian Hox downstream target genes

While the Hox genes were first identified as regulators of anterior–posterior pattern formation during development, it is now clear that Hox genes are also active in normal adult cells and attention is focussed on elucidating a role for Hox proteins in fully differentiated metazoan cells. It is proposed that they will act as key master genes in controlling cellular identity and as such were proposed to regulate genes that are necessary for cell division, cell adhesion and migration, morphological differentiation and apoptosis, in addition to their roles during embryogenesis (Ulijaszek *et al*, 1998).

An array of diverse gene targets has been identified in recent years with roles in cellular identity and function. The experimental procedures used to firstly identify and then to characterise each target has varied greatly, making it difficult to unequivocally state if some are direct targets and even if so, the Hox binding site responsible for the transcriptional regulation may not have been confirmed experimentally. However, utilising the available information a summary of the current likely direct targets is provided in Table 1 and Figure 2.

The first identified mammalian target gene for the Hox transcription factors other than Hox genes themselves was the mouse neural cell adhesion molecule (*N-CAM*) (Jones *et al*, 1992), which mediates adhesion in the

Table 1 Current list of mammalian Hox gene targets

Hox protein	+/−	Target	Species	Reference
Hoxa2	−	Six2	Mouse	Kutejova *et al* (2005)
Hoxa5	+	p53	Mouse	Raman *et al* (2000a)
HOXA5			Human	
HOXA5	+	Progesterone receptor	Human	Raman *et al* (2000b)
HOXA5	+	Pleiotrophin	Human	Chen *et al* (2005)
HOXA5	+	IGFBP-1	Human	Foucher *et al* (2002); Gao *et al* (2002)
HOXA10				
HOXB4				
Hoxa9	−	Osteopontin	Mouse	Shi *et al* (1999, 2001)
Hoxc8				
HOXA9	+	EphB4	Human	Bruhl *et al* (2004)
HOXA10	+	p21	Human	Bromleigh and Freedman (2000)
HOXA10	+	β3-Integrin	Human	Daftary *et al* (2002)
HOXA10	−	EMX2	Human	Troy *et al* (2003)
Hoxa10	+	IGFBP-1	Baboon	Kim *et al* (2003)
Hoxa13, Hoxd13	+	EphA7	Mouse	Salsi and Zappavigna (2006)
HOXB1	+	COL5A2	Human	Penkov *et al* (2000)
Hoxb3	+	TTF-1	Rat	Guazzi *et al* (1994)
Hoxb5	+	SPI3	Mouse	Safaei (1997)
Hoxb5	+	Flk1	Mouse	Wu *et al* (2003)
HOXB7	+	BFGF	Human	Carè *et al* (1996)
Hoxb8	−	N-CAM	Mouse	Jones *et al* (1992)
Hoxb9	+	N-CAM	Mouse	Jones *et al* (1992)
Hoxc8	?	mgl-1	Mouse	Tomotsune *et al* (1993)
Hoxc13	−	Keratins	Mouse	Tkatchenko *et al* (2001)
Hoxd10, b6, b7, b9, c8	+	Renin	Mouse	Pan *et al* (2004)

The list contains likely Hox gene targets and the Hox protein responsible for the *trans*-regulatory effect. The (+/−) symbols represent either a positive or negative regulatory effect on the target gene and (?) symbol indicates an unknown effect. Also, note that although there is some experimental evidence to suggest all are likely direct gene targets, not all have been exclusively verified of being so through *in vivo* experiments. The corresponding reference(s) for each gene target is also shown.
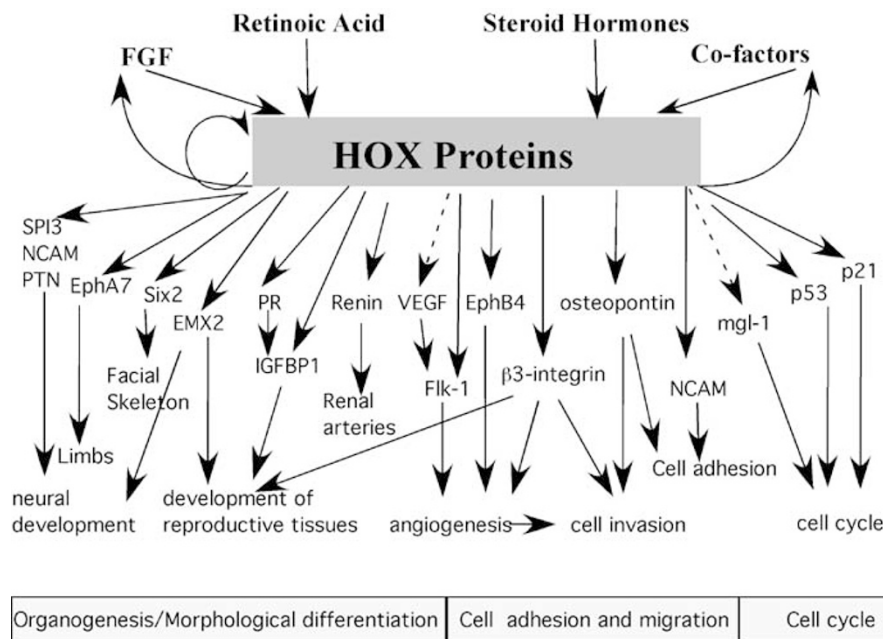
Figure 2 Hox downstream targets are involved in numerous cellular processes including organogenesis, cellular differentiation, cell adhesion and migration, cell cycle and apoptosis. Several downstream targets play multiple roles in several pathways, with many acting as transcription factors and regulating their own subset of genes. Hox proteins also participate in autoregulatory circuits in addition to regulating expression of some binding cofactors and other factors that influence *Hox* expression. Targets not shown on this figure include the differentiation markers alpha-2(V) collagen, the murine hair keratin family and the TTF-1. A dotted line indicates direct regulation has not yet been confirmed.

nervous system and plays a role in neural induction during development. Subsequently, other targets involved in neurogenesis have been identified. The gene encoding SP13, a serine protease inhibitor detected in the mouse nervous system was found regulated by Hoxb5 (Safaei, 1997), although Hox binding sites in the

promoter are yet to be defined. Recently, a homeodomain protein called Six2, which plays a role in the process by which neural crest cells form the facial skeleton was shown to be a downstream target of Hoxa2. Hoxa2 functions during development of the second branchial arch by limiting its formation to a specific region (Grammatopoulos et al, 2000) and as part of this process Hoxa2 negatively regulates Six2 expression (Kutejova et al, 2005). The pleiotrophin (PTN) gene promoter has just recently been characterised as a direct target for HOXA5 and its protein product displays neurite outgrowth-promoting activity (Chen et al, 2005).

Hox targets associated with mammalian reproduction have also been identified. HOXA10 and HOXA11 are expressed in both embryonic and adult reproductive tracts with HOXA10 being essential for uterine development (Block et al, 2000). EMX2 is a divergent homeodomain protein responsible for urogenital tract development and is encoded by a gene negatively regulated by Hoxa10 in endometrial cells (Troy et al, 2003). EMX2 is an orthologue of Drosophila empty spiracles (ems), which is positively regulated by AbdB, the orthologue of HOXA10 (Jones and McGinnis, 1993), although the identified binding sites show no similarity. Both EMX2 and ems also participate in cerebral cortex development (Muzio and Mallamaci, 2003).

HOXA10 and HOXA11 are required for successful embryo implantation and their expression is regulated by sex steroids (Taylor et al, 1998). Insulin-like growth factor binding protein-1 (IGFBP-1) is highly expressed during decidualisation and this expression is progesterone-dependent, as is Hoxa10 (Bell et al, 1991). HOXA10 activates the IGFBP-1 promoter in various types of endometrial cells (Gao et al, 2002) in cooperation with FKHR, a member of the FOXO subfamily of forkhead (fkh) transcription factors (Kim et al, 2003). The IGFBP-1 promoter is also activated by the combined action of the progesterone receptor (PR) and FOXO1A, another fkh transcription factor (Kim et al, 2005). The HOXA5 protein interacts with FKHR to stimulate the IGFBP-1 promoter (Foucher et al, 2002), and interestingly HOXA5 can activate the PR promoter (Raman et al, 2000b), thus revealing some of the intricate pathways and interactions by which Hox proteins regulate embryo implantation.

The above results suggest that multiple HOX proteins may interact with FKHR, thus expanding the list of potential Hox cofactors. The Drosophila forkhead (fkh) gene is upregulated in vivo by scr, a member of the homeotic complex, while Hoxb4 stimulates a direct trans-activation of the TALE class homeodomain gene IRX5 in Xenopus laevis (Theokli et al, 2003). Thus, Hox proteins not only regulate expression of other Hox genes, but also the genes encoding Hox cofactors.

Endothelial cell differentiation is required as a first step of vascular development, including angiogenesis. The earliest marker for endothelial cells is the vascular endothelial growth factor receptor flk1. The gene encoding flk1 is trans-activated by Hoxb5 binding to a site that is located within the first exon, rather than in the promoter region (Wu et al, 2003). Expression of its ligand, VEGF, is upregulated by Hoxb7 (Carè et al, 1996) but is not yet confirmed as a direct target. Additional targets for Hox proteins during angiogenesis have also been identified. The Eph receptor family is the largest known subfamily of receptor tyrosine kinases and a

positive regulation of the EphB4 receptor gene promoter by Hoxa9 was shown to modulate endothelial cell migration and tube formation (Bruhl et al, 2004). In addition, the β3-integrin subunit gene expression is directly regulated by HOXA10 in endometrial cells (Daftary et al, 2002) and provides yet another Hox target involved in both angiogenesis and embryo implantation.

Another member of the Eph family of receptor tyrosine kinases, EphA7, is known to have a role in limb development and the promoter of the mouse gene has now been shown to be directly activated by both Hoxa13 and Hoxd13 (Salsi and Zappavigna, 2006). Interestingly, a mutant version of the HOXD13 protein found in patients with a polydactyl syndrome was unable to activate the EphA7 gene promoter, thus providing a strong connection between the Hox Group 13 family members and limb development by regulating expression of Eph proteins.

Renin is a component of the renin–angiotensin system that regulates blood pressure. Expression of renin is developmentally regulated and is detected in newly formed renal arterial branches (Jones et al, 1990). The murine renin gene promoter (Pan et al, 2004) is activated by Hox/Pbx heterodimers through a single Hox binding site in the gene promoter, providing another link between Hox proteins and blood vessel formation.

Osteopontin (OPN) is a secreted protein that regulates adhesion and migration of cells in a number of biological processes including angiogenesis, inflammation, apoptosis and bone remodelling. Hoxc8 (Shi et al, 1999) and Hoxa9 (Shi et al, 2001) both act as transcriptional repressors by binding to a single Hox binding motif in the Opn promoter. These studies identified another cofactor for Hox proteins as Smad1 formed heterodimers with Hoxc8 and Smad4 formed heterodimers with Hoxa9. The Smad family of proteins mediate signal transduction pathways stimulated by bone morphogenetic proteins (BMP). Smad–Hox heterodimers dislodge Hox monomers from the binding site, which results in BMP driven initiation of transcription from the mouse Opn promoter (Shi et al, 1999, 2001). Analysis of the Opn promoter sequence reveals both Hoxc8 and Hoxa9 utilise the same consensus TAAT motif to mediate the repression.

Downstream HOX targets with roles in the cell cycle are predicted although few examples have been identified. HOXA5 is a potent transcriptional activator of the tumour-suppressor gene p53. A large increase in p53-promoter-dependent reporter activity was observed in breast cancer cell lines upon cotransfection with HOXA5, an effect not seen with other selected HOX members, HOXB4, HOXB5 and HOXB7 (Raman et al, 2000a). The same trans-regulatory effects were observed for the murine p53 homologue, with a single Hox binding element 200 bp upstream from the putative transcription start site identified as responsible for trans-activation.

Another target gene encoding a product with cell cycle function is the cyclin-dependent kinase inhibitor p21, and both negative and positive regulatory effects were observed (Bromleigh and Freedman, 2000). HOXA10, but not HOXA9 or HOXB7, was shown to strongly activate the p21 promoter in both monocytic and fibroblast cell lines. Mutation studies identified a specific binding site responsible for this activation and revealed an adjacent

consensus Hox binding element responsible for a negative regulatory role.

Hox proteins are known to regulate the expression of other Hox genes and, as discussed above, some genes encoding cofactors. A further level of autoregulatory control is provided with the identification of basic fibroblast growth factor (*bFGF*) as a direct target for HOXB7 (Carè *et al*, 1996). Gradients of several molecules including FGF and retinoic acid are responsible for controlling Hox gene expression and segment formation in the embryo as reviewed recently (Diez del Corral and Storey, 2004), and it appears Hox proteins may in turn regulate expression of these gradients.

Finally, expression of *Pax3*, a paired-homeodomain class gene, is upregulated in the murine neural tube by the combined action of POU-domain transcription factors and Hox proteins (Pruitt *et al*, 2004). Brn2 mediates *Pax3* expression through a synergistic input from a subset of Hox proteins, potentially via different binding elements in the *Pax3* promoter. The *trans*-activation of the *Pax3* promoter highlights the complexity of mechanisms by which Hox proteins are involved in directing biological specificity. For this target gene the Hox family cooperate with members from both the TALE and POU family of divergent homeodomain-containing transcription factors to regulate expression of a fourth homeodomain family, the paired-homeodomain class.

Hox binding sites identified in target gene promoters

In order to better understand Hox transcriptional specificity, we examined the identified Hox binding motifs of the mammalian targets in an effort to characterise base pair selections outside of the canonical

TAAT motif. Depicted in Figure 3 are the characterised Hox binding motifs in the promoter regions of target genes as well as flanking sequences. Binding motifs located in the promoter regions of the *bFGF, PR, IGFBP-1, N-CAM*, thyroid transcription factor-1 (*TTF-1*), *EPHB4, β3 Integrin* subunit, *PTN, renin* and *p21* genes result in transcriptional activation. A T(T/A)AT motif was experimentally shown to be responsible for repression by Hox proteins within the promoter regions of *p21, OPN* and *EMX2*. It is therefore possible that base pairs flanking the canonical four-base Hox binding motif play a role in Hox binding specificity.
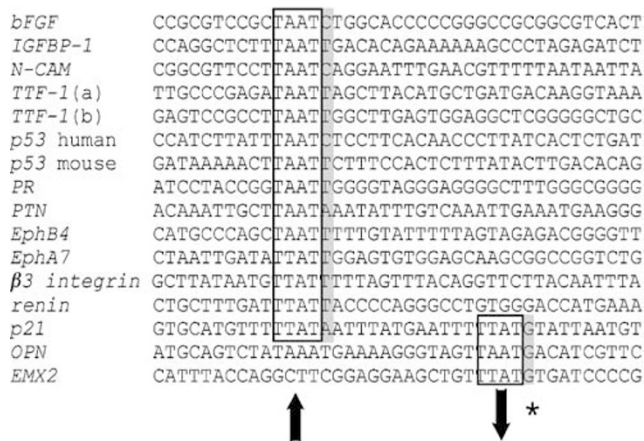
The T(T/A)AT core element and the flanking base pairs all contain a T or C residue at position 5, except the *PTN* and *p21* HOXA10-activated motif, which has an A base at this position. *In vitro* binding site studies also demonstrated a preference for a T at position 5 (Catron *et al*, 1993). The limited number of defined Hox binding sites as shown in Figure 3 makes it as yet impossible to determine if further base selectivity outside the canonical 4 base motif can be ascribed to binding by a particular Hox protein or Hox–cofactor complex.

A paradox of Hox transcriptional specificity is the ability to both suppress and activate gene expression through similar DNA motifs. Of the three T(T/A)AT suppressor motifs, all contain a G at position 5, whereas none of the activator-motifs do. In addition the *Six2* gene promoter contains two binding sites for Hoxa2, TAATTT and TAATGG (Kutejova *et al*, 2005). It is not yet known through which site Hoxa2 acts to repress *Six2* expression but it is of interest that one of the binding sites contains a G at position 5. Although it should be stressed the present sample size is small and therefore no definitive conclusion can be drawn it will be interesting to observe if this trend is upheld when the sample size is expanded by characterising more binding sites where Hox proteins act as repressors.

Figure 3 The identified Hox DNA binding elements of the promoter regions of Hox target genes with flanking sequences. The *TTF-1* gene promoter contains two binding elements and both are shown. The boxed elements indicated with an upwards arrow result in transcriptional activation when bound by the Hox protein, while a downwards arrow indicates repression. The boxed elements further 3′ in the *p21* gene promoter and in *OPN* and *EMX2* have been shown to bind Hox protein complexes with subsequent transcriptional repression. The shaded nucleotides represent the immediate flanking base pair of the canonical Hox element (position 5). Elements where a repressing Hox activity was observed all contain a G at 'position 5' as labelled with an asterisk.

Conclusions

This review has focussed on the transcriptional and functional specificity elicited by the Hox proteins, outlining homeodomain–DNA interactions and the gain of binding specificity through cooperative binding with cofactors. Furthermore, the aim was to summarise the list of known Hox downstream target-genes in mammalian systems. Unfortunately, this list is still very short and contains genes not conclusively shown to be direct targets, albeit strongly suggestive. To fully appreciate the role of Hox proteins both in development and adult tissues, future research inevitably needs to characterise more direct gene targets. This raises the question of why so few targets have yet been identified, and highlights the complexity in which Hox proteins operate in the cellular machinery, potentially rendering the Hox targets to be very elusive.

Obvious problems are that more than one Hox protein can potentially bind to the same element and it is difficult to distinguish between direct and indirect targets. Many Hox targets are also transcription factors, which in turn regulate their own subset of regulatory factors. Yet another dimension of complexity is introduced with the extensive auto- and crossregulation reported for the Hox complex, which was not discussed herein. Similarly, the view that Hox proteins may

regulate gene expression by pathways not involving direct DNA binding further complicates the search for gene targets (Shen *et al*, 2001; Williams *et al*, 2005). Microarray technology has been utilised to identify large numbers of possible gene targets regulated by specific Hox proteins (Lei *et al*, 2005; Williams *et al*, 2005), but further experimentation is still required for verification as direct targets. The use of chromatin immunoprecipitation (ChIP) assays to search for downstream targets would also greatly increase the likelihood that direct targets are identified. Many of the target genes described here have only been shown to *trans*-regulate the respective gene promoters in *in vitro* models, still raising questions regarding their *in vivo* functionality.

The TALE-class homeodomain proteins Pbx and Meis are well-known Hox cofactors, but other proteins can also fulfil such a role *in vivo*. In particular, it would be interesting to further elucidate the cooperative interactions between Hox and Smad proteins, which themselves are essential for proper transforming growth factor-β (TGF-β) signalling. From what is known thus far, Hox proteins interact with numerous regulatory pathways, including FGF, BMP, retinoic acid and sex steroid signalling through cofactor-dependent and -independent regulation. The emergence of FOXO fkh transcription factors as potential cofactors that provide a synergistic input with Hox proteins could explain tissue-specific functionality where simply Hox expression patterns fail to do so.

As another dilemma regarding Hox transcriptional specificity is concerned with the observation that many Hox members act as both transcriptional activators and suppressors depending on cellular context, this review proposes a mechanism in which the flanking nucleotides 3′ of the canonical Hox binding elements could act as an activator-suppressor switch. Taking into account the high sequence homology of relevant peptide sequences within the Hox homeodomains, it is further proposed that the switch is highly dependent on cofactor interactions, which again could account for tissue-specific *trans*-regulatory potential of the Hox complex.

It is clear that in order to further elucidate many of the unresolved issues regarding Hox *in vivo* transcriptional specificity, more direct target genes must be defined. Preferably, new target genes would be identified from animals such as the mouse, which ultimately would allow for more comprehensive studies utilising conditional knockout and targeted knock-in approaches.

Acknowledgements

References

Ades SE, Sauer RT (1994). Differential DNA-binding specificity of the engrailed homeodomain: the role of residue 50. *Biochemistry* **33**: 9187–9194.
Ades SE, Sauer RT (1995). Specificity of minor-groove and major-groove interactions in a homeodomain-DNA complex. *Biochemistry* **34**: 14601–14608.
Bell SC, Jackson JA, Ashmore J, Zhu HH, Tseng L (1991). Regulation of insulin-like growth factor-binding protein-1 synthesis and secretion by progestin and relaxin in long term cultures of human endometrial stromal cells. *J Clin Endocrinol Metab* **72**: 1014–1024.
Block K, Kardana A, Igarashi P, Taylor HS (2000). In utero diethylstilbestrol (DES) exposure alters Hox gene expression in the developing mullerian system. *FASEB J* **14**: 1101–1108.
Bromleigh VC, Freedman LP (2000). p21 is a transcriptional target of HOXA10 in differentiating myelomonocytic cells. *Genes Dev* **14**: 2581–2586.
Bruhl T, Urbich C, Aicher D, Acker-Palmer A, Zeiher AM, Dimmeler S (2004). Homeobox A9 transcriptionally regulates the EphB4 receptor to modulate endothelial cell migration and tube formation. *Circ Res* **94**: 743–751.
Carè A, Silvani A, Meccia E, Mattia G, Stoppacciaro A, Parmiani G et al (1996). HOXB7 constitutively activates basic fibroblast growth factor in melanomas. *Mol Cell Biol* **16**: 4842–4851.
Catron KM, Iler N, Abate C (1993). Nucleotides flanking a conserved TAAT core dictate the DNA binding specificity of three murine homeodomain proteins. *Mol Cell Biol* **13**: 2354–2365.
Chang C-P, Shen W-F, Rozenfeld S, Lawrence HJ, Largman C, Cleary ML (1995). Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins. *Genes Dev* **9**: 663–674.
Chen H, Rubin E, Zhang H, Chung S, Jie CC, Garrett E et al (2005). Identification of transcriptional targets of HOXA5. *J Biol Chem* **280**: 19373–19380.
Daftary GS, Troy PJ, Bagot CN, Young SL, Taylor HS (2002). Direct regulation of beta-3-integrin subunit gene expression by HOXA10 in endometrial cells. *Mol Endocrinol* **16**: 571–579.
Desplan C, Theis J, O'Farrel PH (1988). The sequence specificity of homeodomain-DNA interaction. *Cell* **54**: 1081–1090.
Diez del Corral R, Storey KG (2004). Opposing FGF and retinoid pathways: a signalling switch that controls differentiation and patterning onset in the extending vertebrate body axis. *BioEssays* **26**: 857–869.
Ekker SC, von Kessler DP, Beachy PA (1992). Differential DNA sequence recognition is a determinant of specificity in homeotic gene action. *EMBO J* **11**: 4059–4072.
Foucher I, Volovitch M, Frain M, Kim JJ, Souberbielle J-C, Gan L et al (2002). Hoxa5 overexpression correlates with IGFBP1 upregulation and postnatal dwarfism: evidence for an interaction between Hoxa5 and forkhead box transcription factors. *Development* **129**: 4065–4074.
Gao J, Mazella J, Tseng L (2002). Hox proteins activate the IGFBP-1 promoter and supress the function of hPR in human endometrial cells. *DNA Cell Biol* **21**: 819–825.
Gebelein B, Culi J, Ryoo HD, Zhang W, Mann RS (2002). Specificity of *distalless* repression and limb primordia development by abdominal Hox proteins. *Dev Cell* **3**: 487–498.
Grammatopoulos GA, Bell E, Toole L, Lumsden A, Tucker AS (2000). Homeotic transformation of branchial arch identity after Hoxa2 overexpression. *Development* **127**: 5355–5365.
Greer JM, Puetz J, Thomas KR, Capecchi MR (2000). Maintenance of functional equivalence during paralogous Hox gene evolution. *Nature* **403**: 661–664.
Guazzi S, Lonigro R, Pintonello L, Boncinelli E, Di Lauro R, Mavilio F (1994). The thyroid transcription factor-1 gene is a candidate target for regulation by Hox proteins. *EMBO J* **13**: 3339–3347.
Jones B, McGinnis W (1993). The regulation of empty spiracles by abdominal-B mediates an abdominal segment identity function. *Genes Dev* **7**: 229–240.
Jones CA, Sigmund CD, McGowan RA, Kane-Haas CM, Gross KW (1990). Expression of murine renin genes during fetal development. *Mol Endocrinol* **4**: 375–383.
Jones FS, Prediger EA, Bittner DA, DeRobertis EM (1992). Cell adhesion molecules as targets for Hox genes: neural cell adhesion molecule promoter activity is modulated by

cotransfection with Hox-2.5 and -2.4. *Proc Natl Acad Sci USA* **89**: 2086–2090.

Kim JJ, Buzzio OL, Li S, Lu Z (2005). Role of FOXO1A in the regulation of insulin like growth factor binding protein-1 in human endometrial cells: interaction with progesterone receptor. *Biol Reprod* **73**: 833–839.

Kim JJ, Taylor HS, Akbas GE, Foucher I, Trembleau A, Jaffe RC et al (2003). Regulation of insulin-like growth factor binding protein-1 promoter activity by FKHR and HOXA10 in primate endometrial cells. *Biol Reprod* **68**: 24–30.

Kissinger CR, Liu B, Martin-Blanco E, Kornberg TB, Pabo CO (1990). Crystal structure of an engrailed homeodomain-DNA complex at 2.8A resolution: a framework for understanding homeodomain-DNA interactions. *Cell* **63**: 579–590.

Kutejova E, Engist B, Mallo M, Kanzler B, Bobola N (2005). *Hoxa2* downregulates *Six2* in the neural crest-derived mesenchyme. *Development* **132**: 469–478.

Kuziora MA, McGinnis W (1989). A homeodomain substitution changes the regulatory specificity of the *deformed* protein in *Drosophila* embryos. *Cell* **59**: 563–571.

LaRonde-LeBlanc NA, Wolberger C (2003). Structure of HoxA9 and Pbx1 bound to DNA: Hox hexapeptide and DNA recognition anterior to posterior. *Genes Dev* **17**: 2060–2072.

Lei H, Wang H, Juan AH, Ruddle FH (2005). The identification of *Hoxc8* target genes. *Proc Natl Acad Sci USA* **102**: 2420–2424.

Levine M, Hoey T (1988). Homeobox proteins as sequence-specific transcription factors. *Cell* **55**: 537–540.

Lewis EB (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* **276**: 565–570.

Lin L, McGinnis W (1992). Mapping functional specificity in the Dfd and Ubx homeo domains. *Genes Dev* **6**: 1071–1081.

Malicki J, Schughart K, McGinnis W (1990). Mouse *Hox-2.2* specifies thoracic segmental identity in *Drosophila* embryos and larvae. *Cell* **63**: 961–967.

Mann RS, Morata G (2000). The developmental and molecular biology of genes that subdivide the body of *Drosophila*. *Annu Rev Cell Dev Biol* **16**: 243–271.

McGinnis N, Kuziora MA, McGinnis W (1990). Human *Hox-4.2* and *Drosophila deformed* encode similar regulatory specificities in *Drosophila* embryos and larvae. *Cell* **63**: 969–976.

McGinnis W, Krumlauf R (1992). Homeobox genes and axial patterning. *Cell* **68**: 283–302.

Medina-Martinez O, Ramirez-Solis R (2003). *In vivo* mutagenesis of the *Hoxb8* hexapeptide domain leads to dominant homeotic transformations that mimic the loss-of-function mutations in genes of the *Hoxb* cluster. *Dev Biol* **264**: 77–90.

Muzio L, Mallamaci A (2003). Emx1, emx2 and pax6 in specification, regionalization and arealization of the cerebral cortex. *Cereb Cortex* **13**: 641–647.

Otting G, Qian YQ, Billeter M, Muller M, Affolter M, Gehring WJ et al (1990). Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution. *EMBO J* **9**: 3085–3092.

Pan L, Jones CA, Glenn ST, Gross KW (2004). Identification of a novel region in the proximal promoter of the mouse renin gene critical for expression. *Am J Physiol* **286**: F1107–F1115.

Pellerin I, Schnabel C, Catron KM, Abate C (1994). Hox proteins have different affinities for a consensus DNA site that correlate with the positions of their genes on the *hox* cluster. *Mol Cell Biol* **14**: 4532–4545.

Penkov D, Tanaka S, DiRocco G, Berthelsen J, Blasi F, Ramirez F (2000). Cooperative interactions between PBX, PREP, and HOX proteins modulate the activity of the alpha-2(V) collagen (COL5A2) promoter. *J Biol Chem* **275**: 16681–16689.

Phelan ML, Rambaldi I, Featherstone MS (1995). Cooperative interactions between HOX and PBX proteins mediated by a conserved peptide motif. *Mol Cell Biol* **15**: 3989–3997.

Pinsonneault J, Florence B, Vaessin H, McGinnis W (1997). A model for *extradenticle* function as a switch that changes HOX proteins from repressors to activators. *EMBO J* **16**: 2032–2042.

Piper DE, Batchelor AH, Chang C-P, Cleary ML, Wolberger C (1999). Structure of a HoxB1-Pbx1 heterodimer bound to DNA: Role of the hexapeptide and a fourth homeodomain helix in complex formation. *Cell* **96**: 587–597.

Pruitt SC, Bussman A, Maslov AY, Natoli TA, Heinaman R (2004). Hox/Pbx and Brn binding sites mediate *Pax3* expression *in vitro* and *in vivo*. *Gene Expr Patterns* **4**: 671–685.

Qian YQ, Billeter M, Otting G, Muller M, Gehring WJ, Wuthrich K (1989). The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. *Cell* **59**: 573–580.

Raman V, Martensen SA, Reisman D, Evron E, Odenwald WF, Jaffee E et al (2000a). Compromised HOXA5 function can limit p53 expression in human breast tumours. *Nature* **405**: 974–978.

Raman V, Tamori A, Vali M, Zeller K, Korz D, Sukumar S (2000b). HOXA5 regulates expression of the progesterone receptor. *J Biol Chem* **275**: 26551–26555.

Safaei R (1997). A target of the HoxB5 gene from the mouse nervous system. *Dev Brain Res* **100**: 5–12.

Salsi V, Zappavigna V (2006). Hoxd13 and Hoxa13 directly control the expression of the EphA7 ephrin tyrosine kinase receptor in developing limbs. *J Biol Chem* **281**: 1992–1999.

Scott MP (1992). Vertebrate homeobox gene nomenclature. *Cell* **71**: 551–553.

Seeman NC, Rosenberg JM, Rich A (1976). Sequence-specific recognition of double helical nucleic acids by proteins. *Proc Natl Acad Sci USA* **73**: 804–808.

Shen W, Chrobak D, Krishnan K, Lawrence HJ, Largman C (2004). HOXB6 protein is bound to CREB-binding protein and represses globin expression in a DNA binding-dependent, PBX interaction-independent process. *J Biol Chem* **279**: 39895–39904.

Shen WF, Chang C-P, Rozenfeld S, Sauvageau G, Humphries RK, Lu M et al (1996). Hox homeodomain proteins exhibit selective complex stabilities with Pbx and DNA. *Nucleic Acids Res* **24**: 898–906.

Shen WF, Krishnan K, Lawrence HJ, Largman C (2001). The HOX homeodomain proteins block CBP histone acetyltransferase activity. *Mol Cell Biol* **21**: 7509–7522.

Shen WF, Montgomery JC, Rozenfeld S, Moskow JJ, Lawrence HJ, Buchberg AM et al (1997a). AbdB-like Hox proteins stabilize DNA binding by the Meis1 homeodomain proteins. *Mol Cell Biol* **17**: 6448–6458.

Shen WF, Rozenfeld S, Lawrence HJ, Largman C (1997b). The Abd-B-like Hox homeodomain proteins can be subdivided by the ability to form complexes with Pbx1a on a novel DNA target. *J Biol Chem* **272**: 8198–8206.

Shi X, Bai S, Li L, Cao X (2001). Hoxa-9 represses transforming growth factor-beta-induced osteopontin gene transcription. *J Biol Chem* **276**: 850–855.

Shi X, Yang X, Chen D, Chang Z, Cao X (1999). Smad1 interacts with homeobox DNA-binding proteins in bone morphogenetic protein signaling. *J Biol Chem* **274**: 13711–13717.

Taylor HS, Arici A, Olive D, Igarashi P (1998). HOXA10 is expressed in response to sex steroids at the time of implantation in the human endometrium. *J Clin Invest* **101**: 1379–1384.

Theokli C, El-Kadi ASM, Morgan R (2003). TALE class homeodomain gene *Irx5* is an immediate downstream target for *Hoxb4* transcriptional regulation. *Dev Dyn* **227**: 48–55.

Tkatchenko AV, Visconti RP, Shang L, Papenbrock T, Pruett ND, Ito T et al (2001). Overexpression of Hoxc13 in differentiating keratinocytes results in downregulation of a novel hair keratin gene cluster and alopecia. *Development* **128**: 1547–1558.

Tomotsune D, Shoji H, Wakamatsu Y, Kondoh H, Takahashi N (1993). A mouse homologue of the *Drosophila* tumour-suppressor gene *I(2)gI* controlled by Hox-C8 *in vivo*. *Nature* **365**: 69–72.

96

Troy PJ, Daftary GS, Bagot CN, Taylor HS (2003). Transcriptional repression of peri-implantation EMX2 expression in mammalian reproduction by HOXA10. *Mol Cell Biol* **23**: 1–13.

Ulijaszek SJ, Johnston FE, Preece MA (1998). *The Cambridge Encyclopedia of Human Growth and Development*. Cambridge University Press: Cambridge.

Walsh B (2003). Population-genetic models of the fates of duplicate genes. *Genetica* **118**: 279–294.

Williams TM, Williams ME, Kuick R, Misek D, McDonagh K, Hanash S *et al* (2005). Candiate downstream regulated genes of HOX group 13 transcription factors with and without monomeric DNA binding capability. *Dev Biol* **279**: 462–480.

Wu Y, Moser M, Bautch VL, Patterson C (2003). HoxB5 is an upstream transcriptional switch for differentiation of the vascular endothelium from precursor cells. *Mol Cell Biol* **23**: 5680–5691.